

# コールセンターにおける音声技術の利用 ～音声のパラ言語情報・非言語情報を含めた活用～

桜井淳宏<sup>†</sup> 木村晋太<sup>†</sup>

コールセンターは音声によるコミュニケーションの重要な活動領域であるため、音声情報処理に対する期待は特に大きい。また、コンプライアンスの維持およびサービス品質の向上のために通話録音は必須になってきている。その結果、膨大な音声データが蓄積されるようになり、さらにそれを活用する動きが活発化している。一方、実業務において音声技術を活用するには、現場の経験に基づいたきめ細かい配慮や工夫が必要である。当社は通話録音・音声認識・対話分析・非言語やパラ言語情報の分析といった技術をコールセンターの業務改善に応用し、音声技術の活用を広げようとしている。本発表では、事例を紹介しながら、コールセンター業界に必要とされる音声技術の具体像を探りたい。

## The use of speech technologies in call centers - Including para- and non-linguistic information

ATSUHIRO SAKURAI<sup>†1</sup> SHINTA KIMURA<sup>†</sup>

Call centers constitute an important arena for the use of speech processing technology. Moreover, an increasing number of companies perform voice logging, resulting in the accumulation of a large amount of speech data, paving the way for new applications of such resources. On the other hand, the use of speech technology in practice is not an easy task, requiring careful attention and know-how. Our company specializes in the use of speech technologies ranging from voice logging, speech recognition, dialog analysis, and para- or non-linguistic analysis to create new solutions for the call center industry. In this report, we show some examples and try to grasp what is really expected from speech technologies in the call center industry.

### 1. はじめに

コールセンターでは音声是最も重要なコミュニケーション手段であるため、音声情報処理技術（以下：音声技術）を導入するメリットは大きいとされている。コールセンターにおける音声技術といえば、音声認識と音声合成を組み合わせた音声自動応答システム（IVR: interactive voice response system）を連想しやすいが、既存の IVR システムにおける音声認識の利用は高度な対応を必要としない業務に限定される。また、音声合成は固定文のオフライン生成という利用方法が大勢をしめる。実際、音声技術と自然言語処理を組み合わせた対話システムがコールセンター業界で普及するのにまだ時間がかかると思われる。

しかしながら、音声技術の活用は IVR 以外に様々な分野に及ぶ。IVR は人間対マシンのリアルタイムインタラクションを実現しようとしているが、人間対人間、もしくは非リアルタイム（オフライン）分析といった領域にも当然ながら音声技術を利用できる。また、音声合成や音声認識以外にもパラ言語・非言語情報の利用も可能である。

ただし、実用的なソリューションを構築するにあたり、精度が不完全であるという音声技術の本質的ともいえるべき限界を理解し、機械の持つ特長（速い、疲れない）を最大

限に活かすことが鍵となる。また、どのようなソリューションが適切であるか見極めるため、コールセンターという業界の現状や課題を把握する必要がある。本発表では、コールセンターの現状や課題を把握した上、音声技術の活用の具体例を紹介しながら今後の展開について考える。

### 2. コールセンターの現状や課題

コールセンターは顧客への電話対応業務を行う組織である。その役割は業種によって異なるが、基本的に以下の機能を持つことが多い。

- 事務手続きなどの受付や処理
- 顧客からのクレーム・要望・問い合わせなどの受付やそれに対する説明など
- クロスセルなど能動的なセールスやマーケティング活動

企業活動におけるコールセンターの役割は極めて大きいことはいまでもない。コールセンターは企業にとって顧客と接する貴重なチャンネルであり、企業の「顔」そのものといえる[a]。また、近年、以下の事情によりコールセンターの重要性はますます高まりつつある。

- 法的要因（金融界の場合は金融商品取引法で定められる金融商品に関する説明義務）により、通話録音を導

<sup>†</sup> 株式会社アニモ：http://www.animo.co.jp  
Animo Limited: http://www.animo.co.jp/EN/

a) 最近では e-mail や SNS など新しいタイプのインタラクションが普及している。そこで、テキストによるコミュニケーションにも対応する組織として「コンタクトセンター」という呼称が定着しつつある。

入する企業が多くなってきている。ストレージの価格が下落傾向にあることもそれを後押しする。

- 通話録音の普及に伴い、大量の通話録音データが蓄積されるようになった。そのため、マーケティング情報として録音データを活用しようという動きがある。

コールセンターには1) コスト削減、2) サービス向上、3) 更なる付加価値の追求という基本的な目標がかかげられるが、上記の機能や事情を背景に、目標を達成するための課題として以下の項目があげられる。

- ① オペレータ作業の効率化
- ② オペレータの評価・教育
- ③ 聴取・モニタリング作業の効率化
- ④ 通話データの属性付与・分類
- ⑤ 通話内容の分析・活用など

次節以降、上記の各目的において音声技術をどのように活用できるか、事例を紹介しながら説明する。

### 3. 音声技術の使用例

#### 3.1 オペレータ作業の効率化

専門的な説明が必要な製品を数多く取り扱う企業のコールセンターではオペレータが自らマニュアル類を検索し、それに基づいて顧客に対する説明を行うが、検索にかかる時間はオーバーヘッドとなるため、なるべく短縮する必要がある。また、タイピング音が顧客に聞こえてしまうと悪い印象を与えるだけでなく、会話の流れを妨げるといった問題がある。そこで、オペレータによるマニュアル類の検索作業を支援するため、音声認識技術を利用できる。当社は特定キーワードにマニュアル類をリンクさせ、ワードスポッティング型の音声認識エンジンを利用して検索作業を支援するソフトウェアを製品化している[b]。これを用いて、オペレータは対応案件に関連するキーワードを復唱することで、会話の流れを途切れずにドキュメントの検索を行うことができる。また、手動入力の必要性が減り、タイピング音の問題も軽減される。表1にシステムのコセプトを示す。一つの関連キーワードを発声すると複数のドキュメントがハイライトされ、追加のキーワードでデータをさらに絞る。

#### 3.2 オペレータの評価・教育

オペレータの評価や教育はコールセンターにおいて重要な課題である。現在、スーパーバイザー（上司）による主観評価や擬似顧客（評価者）による「ミステリーコール」という評価手法が主流であるが、より客観的な計量基準群(KPI: key performance indicator)[1]も必要とされる。

表1 オペレータ支援システム「KnowledgeNavi」のコンセプト。

「デジタルカメラ」と「寒い」というキーワードの組み合わせで「デジタルカメラの寒冷地での使用」というマニュアルがポップアップする。

Table 1 KnowledgeNavi agent support system concept. Keywords “digital camera” and “cold” within the conversation lead to the manual “Using the digital camera in cold areas”.

	機能1	機能2	寒冷地	機能3
製品A				
製品B				
デジタルカメラ				
製品C				
...				

オペレータの対応の様子を複数の数値で表現できれば、具体的な基準値に基づく評価尺度の設定が可能になる。運営方法として、従来の主観評価を裏付けるための基本データとして自動で算出した数値を用いることができる。

当社はかねてから対話の流れを表す数値に基づく評価尺度を提供するサービスを展開してきたが、最近では自社の通話録音システムに評価指標用の数値の算出を行うプラグインの提供を始めた[c]。

一般的に、オペレータの評価に有効な指標として以下に示す通り、発声内容（言語情報）に基づく指標とそれ以外の情報（対話及びパラ言語・非言語情報）に基づく指標がある。以下、それぞれについて述べる。

#### (1) 発声内容に基づくオペレータ評価指標

発声内容に基づく評価指標として、通話の各段階（オープニング、お礼／お詫び、問診、本人確認、クロージング）に言うべき言葉（OK ワード）を発声したか、または禁じられた言葉や表現（NG ワード）を発声していないかなどといったチェック項目がある。このようなタスクには音声認識技術を活用できるが、当社ではキーワードの検索に特化したワードスポッティング型音声認識ソフトウェア[d]でそれを実現する。以下の事例は、ある住宅設備メーカーのコールセンターのオペレータを対象に行った発言キーワードの調査である。オペレータの勤続年数に応じてベテラン・中堅・新人に分けて調査を行った。図1はOK ワード／NG ワードの調査結果の一部である。現場でのヒアリングの後、下記のような結論に至った。

- 新人：言葉使いの教育は比較的よくできている
- ベテラン：OK ワード発言の基準を満たしているが、NG ワードの発声が多い（慣れから来る略語の利用など）

b) リアルタイムナレッジ検索システム KnowledgeNavi

c) オペレータ評価システム VoiceTracking Quality Manager

d) キーワード検索システム VoiceTracking KeywordFinder

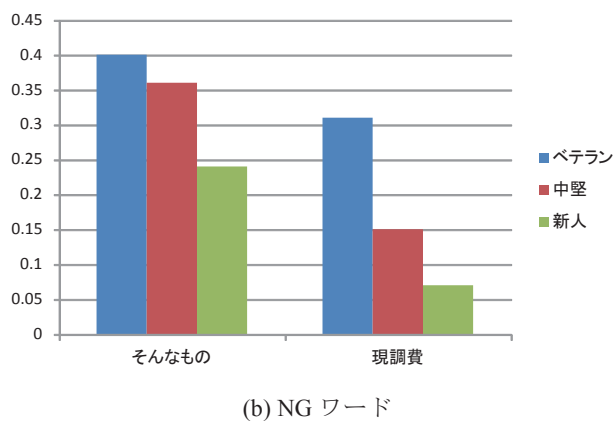
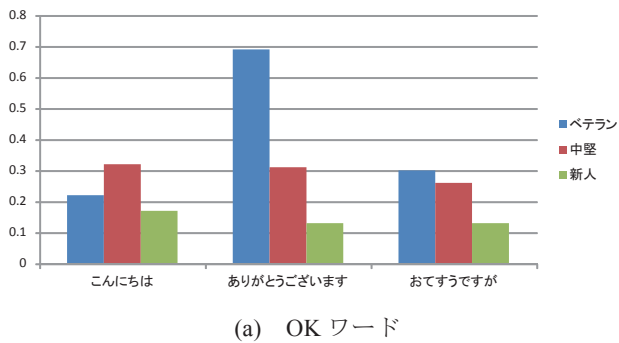


図 1 キーワードの発言によるオペレータ評価。  
 縦軸は通話あたりの平均発話回数である。

Figure 1 Agent evaluation based on keyword utterance.

The vertical axis represents the average number of utterances per call.

(2) 対話及びパラ言語・非言語情報に基づくオペレータ評価指標

多くの通話録音システムは通話の上り・下りを別々のチャンネルで録音するステレオ録音方式を採用している。このようなシステムを利用すればオペレータと顧客の発声を分離して分析・評価できる。それぞれの発声のタイミングがわかれば、オペレータの評価に役立つ様々な指標が得られる。表 2 にはオペレータ評価で利用されている主な分析パラメータを示す。

図 2 に前記の住宅設備メーカーのコールセンターの分析結果を示す。ここも同じく結果をベテラン・中堅・新人オペレータに分けた。このデータは有意差に基づく厳密な結論を導くためのものではないが、主観による知見を説明するための裏付けデータとして利用できる。例えば、新人オペレータは短時間で通話を終わらせようとするが、経験を積むにつれ、顧客と会話できるようになることや新人の平均無音割合及び「かぶり」の回数が高いことなどは現場の常識と一致する結果である。

表 2 オペレータ評価指標として利用される対話分析パラメータ

Table 2 Dialog analysis parameters used for evaluating operator's performance

対話分析パラメータ	意味
発言量	オペレータ・顧客それぞれの発言量及びその割合など
かぶりの回数・頻度	発言中に割り込む行為（かぶり）の数値化
あいづちの回数・頻度	「聞き上手」の指標
話速	顧客に合わせた発話スピードになっているか

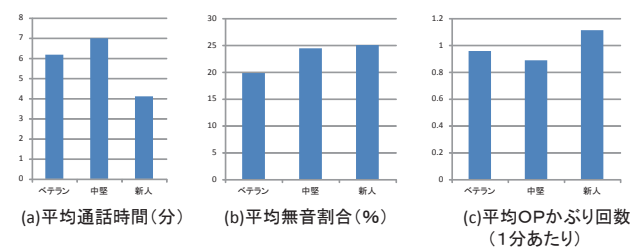


図 2 対話分析によるオペレータの評価

Figure 2 Agent evaluation based on dialog analysis.

3.3 聴取・モニタリング作業の効率化

録音された音声聴取し、通話内容を確認することはコールセンターで日常的に行われる作業である。作業の効率化をはかるために様々な音声技術が利用されるが、ここで代表的な例をあげる。

(1) 話速変換

人手で通話の内容を聴取する場合、本来は少なくとも通話時間と同等の時間が必要である。しかしながら、話速変換技術（基本周波数やスペクトル包絡を保ったまま、時間スケールを変更するための技術）を利用すれば作業時間を短縮できる。話速変換の方法として様々な手法が提案されているが、当社では独自のアルゴリズムを利用している。

(2) 音声認識

通話内容を予め大まかに把握できれば、聴取作業を大幅に効率化できる可能性があるが、そういった用途には音声認識技術は有効である。例えばワードスポッティング型の音声認識処理を施し、重要なキーワードにしおり（タグ）を付与すれば、聴取作業の効率化をはかれる。そのような機能は当社の通話録音ソフトウェアのプラグインとして提供している。

### (3) 聴取対象データの絞り込み

聴取・モニタリングの対象となる通話データそのものを事前に選別することができれば作業の効率化につながる。これは通話データに何らかの属性を付与し、それに基づいて分類することに等しいが、属性付与や分類の用途は聴取作業の効率化に限らず、マーケティング調査や企業コンプライアンス管理など様々な分野に適用できるため、次節で取り上げる。

### 3.4 通話データの属性付与・分類

大規模のコールセンターではすべての録音データを聴取・モニタリングすることは難しい。聴取やモニタリングの対象となる通話を何らかの条件で選別しなければ、無作為にサンプリングすることになってしまう。また、前述の通り、通話に属性を付与することで音声データの事後活用の可能性が広がる。

通話データの属性付与・分類のひとつの方法として、通話音声に音声認識処理を施し、ターゲットとなる特徴を示す内容の有無で属性を決定することが考えられる。このとき、大語彙連続音声認識と高度な意味理解を実現する自然言語処理エンジンの組み合わせが長期的に最も複雑なタスクに対応できると思われる。しかしながら、現在のところ、特に計算機資源が有限の場合、特定キーワードのみを検出するワードスポッティング型音声認識もしくはワードスポッティングと連続音声認識の組み合わせが現実的なソリューションである。

一方、音声には言語情報のほかに話者性など非言語情報やパラ言語情報も含まれるが、それに基づく属性付与や分類も有効である場合もある。以下、そのような技術について説明する。

#### (1) 話者認識による通話データの分類

話者性は通話データの重要な属性のひとつである。コールセンターにおいては、特定話者の通話を検索するという利用シーンが考えられる。そのとき、話者に関する情報が容易に入手できない場合、話者認識技術が有効である。このニーズを満たすのは発声内容に依存しないテキスト非依存型の話者認識技術である[2]。当社は独自の話者認識ソフトウェア[e]をミドルウェアとして販売しているが、今後はクレマーの特定やいたずら電話の防止のほか、マーケティングツールとしての利用が期待される。

#### (2) 声のテンションの分析

通話データの属性付与・分類が必要となるもうひとつの具体例は、トラブル通話（いわゆる「問題呼」）の検出というモニタリング業務である。不満を持つ顧客がコールセンターのオペレータに対して声を荒げるといった場面が典型的な例である。事前にすべての通話を自動で分析し、検出の確率の高い順にデータを配置できれば手動モニタリング

（聴取作業）の効率化をはかれる。

前述の通り、音声には言語情報のほかにパラ言語・非言語情報が含まれるが、「問題呼」という現象は感情と深く関わっているため、パラ言語・非言語情報の影響は大きいといえる。その影響は韻律的特徴や声質など音響的特徴の変化として表れるが、感情はあくまで内なる概念であるため、音響的特徴量との関係を推定・定式化することは難しい。そのため、当社では感情の推定を試みるのではなく、声の物理的特徴量（韻律的特徴およびスペクトル情報）から導出しやすい「声のテンション」という計量値を定義する。この値はあくまで声の状態を表すものであり、感情を直接指すものではない。

テンションの値は音声区間単位で算出するが、対話全体におけるテンションを時系列（以下、テンションパターン）として表す。対話を特徴づけるものとしてテンションパターンを眺めたとき、テキストに表れない対話の動的な特徴が明らかになる。

図3にはオペレータと顧客の対話のテンションパターンの例を示す。それぞれの発声を音声区間に分割した後、物理的特徴量を求める。その後、統計モデルを用いてテンションの値を計算し、時系列（テンションパターン）を求める（図中、オペレータと顧客のパターンを異なる色で表示している）。このテンションパターンは典型的な問題呼（クレーム）である。会話の後半あたりから顧客が怒りを示し、それに対しオペレータは小さくうなずくが、テンションパターンにはそのような様子が表れる。

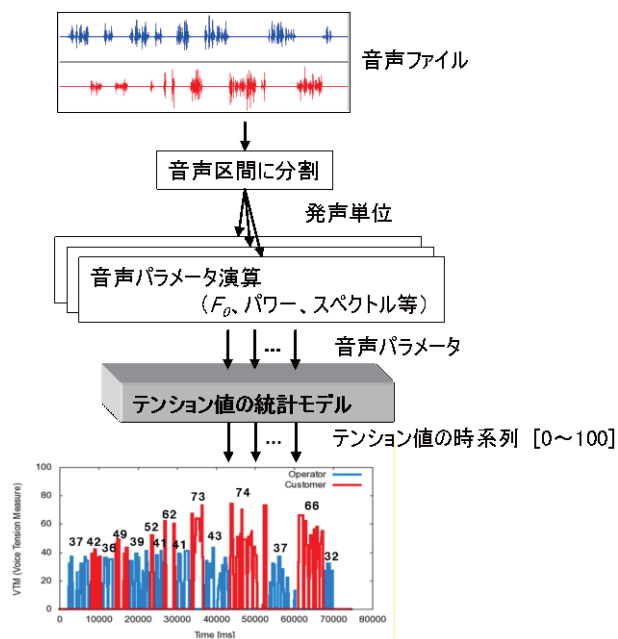


図3 声のテンションの分析  
 Figure 3 Voice tension analysis

e) 話者認識システム VoiceSyncII



このタイプの通話の特徴を捉え、問題呼の検出を可能にするため、「問題呼らしさ」を表す「テンションスコア」という統計量（もしくは ad-hoc 値）を定義する。これに基づいて通話データを分類し、問題呼の検出作業の効率化をはかる。

以下、テンション分析に基づく問題呼検出の事例を示す。本事例は金融機関のコールセンターで行った実証実験である。事業の一環として、債務者へ電話をかけ、状況を確認するという業務があるが、会話がスムーズに行われたかが重要なポイントである。そのため、すべての通話を録音し、手でモニタリング作業を行っている。しかしながら、モニタリング作業は数十名の体制で行っているため、毎日発生する数千単位の通話をすべてチェックすることはできない。そこで、モニタリング作業の効率化が最終的な目的である。

実験ではコールセンターの約 8000 個の通話データの中、何件の「問題呼」を検出できるか検証した。問題呼の実際の件数は未知であるが、通例の比率では 1, 2 件ほど存在するという前提であった。音声データの録音は顧客とオペレータの声を分離して録音できる当社の通話録音システム VoiceTracking で行った。

データセットのすべての通話データのテンションパターンを求め、テンションの平均値及び実験的に調整したしきい値を超える回数に基づく数式を「テンションスコア」とした。このスコアを用いて、図 4 のように通話データをスコア順に並べた。スコアは最高値で正規化した。その後、

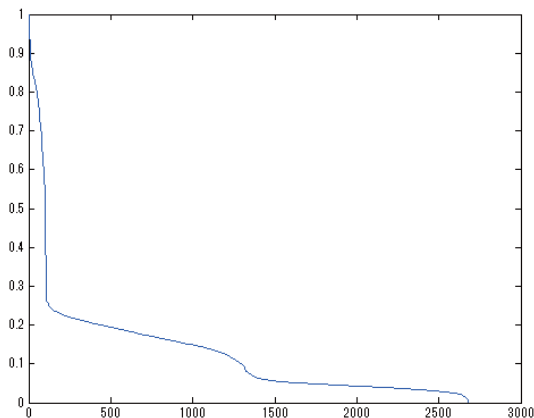


図 4 正規化テンションスコアの順に通話を並べた様子。スコアの高い通話から聴取確認をする。

Figure 4 Calls arranged in descending order of tension scores. Monitoring is done starting from the calls that yielded the highest scores.

スコアの高い通話から実際に聴取確認を行ったところ、スコアのトップ 50 通話のうち、4 件の問題呼が確認された。全通話データの 0.7%に相当する通話数を調べるだけで多

く問題呼の検出が可能であったことになる。そのため、予めテンションパターンに基づく分析を行えば検出作業の効率化につながる事が分かった。

### 3.5 通話内容の分析・活用

前述の通り、通話音声には重要な情報が含まれているため、マーケティングや企業コンプライアンス管理など様々な分野で利用できる。ただし、通話内容をテキストに変換し、さらに意味を理解して要約するという一連のプロセスをすべて自動で行わない限り、音声認識や自然言語処理のポテンシャルをフルに活用できたとはいえない。実際、上記の技術は発展途上にあるため、有効なソリューションにするためには人間の介在が必要になり、作業の効率化をはかることが価値を生み出すという観点が最も現実的である。そこで、費用対効果の側面も含め現状の技術でどのようなソリューションが実現可能であるか探る。

#### (1) 簡易的なトレンド分析

通話内容の簡単な活用法として考えられるのは、世間の新しいキーワードや自社で重要視される用語（社名、製品名など）が通話中にどの頻度で出現するかを分析することである。頻度の時系列の統計分析を行うことによって、有効な関係性を導くことが可能となる。

例えば、特定キーワードにおける曜日別、時間帯別の出現頻度を調べたいとする（図 5）。このような分析は瞬時に結果出力が求められるため、音声から予め情報を抽出しなければならない。当社は、このような用途に向け、新語などにも対応できる高速の検索エンジンを提供している [f]。

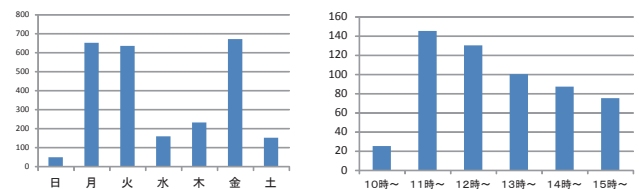


図 5 曜日別・時間帯別のキーワードの出現数  
 Figure 5 Number of keyword utterance by day of the week and by time frame.

#### (2) 音声テキスト変換

通話データの発声内容をテキストに変換できればマーケティングやコンプライアンス管理など様々な分野で長期的な恩恵もたらされることは共通の認識である。ただし、今のところすべての通話内容をテキストに書き起こすのに大きなコストがかかるため、戸惑うユーザーが多い。

現在、導入コストを抑えながら音声テキスト変換技術を導入し、新しい活用方法を開発したいユーザーには、フロントエンドプロセッサとしてワードスポッティング型音声認識を利用することで対象とするターゲットを絞り、重要

f) 高速音素型検索・調査システム VoiceTracking VOC Analyzer

とされる通話のみに対して大語彙連続音声認識を行うという選択肢がある (ハイブリッドタイプの音声認識[3])。

【通話検索確認 画面イメージ】

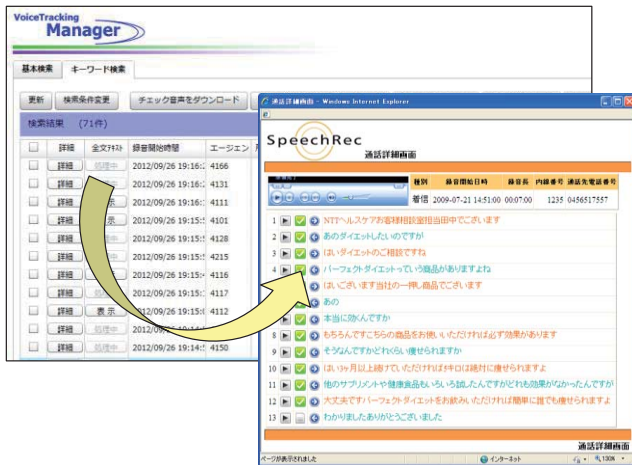


図 6 ワードスポッティング[g]と大語彙連続音声認識システム[h]で構成される「ハイブリッド型音声認識システム」  
Figure 6 “Hybrid” speech recognition system formed by a word-spotting front end and a large-vocabulary continuous speech recognition system.

#### 4. おわりに

コールセンターにおける通話録音の浸透が進む中、音声技術の活用はますます広がりつつある。また、従来考えられてきた音声認識と音声合成のほか、パラ言語・非言語情報の活用も可能であることがわかった。今後はより幅広い用途での音声技術の活用をめざすとともに、音声技術がコールセンターのITシステムの欠かせない部分になることを願っている。

#### 参考文献

- 1) <http://ja.wikipedia.org/wiki/重要業績評価指標>
- 2) 越仲孝文, 篠田浩一, “話者認識の国際動向,” 日本音響学会誌 69 巻 7 号, 2013.
- 3) “製品トレンド/音声認識：期待値と活用レベルにギャップ,” Computer Telephony, 9 月号, 2013.

g) キーワード検索システム VoiceTracking KeywordFinder  
h) 高精度音声認識 (ASR) ソリューション SpeechRec (NTT-IT 社)